

Stolle

PAGE: 1

**RAW SEQUENCE LISTING  
PATENT APPLICATION US/08/851,089**

1652 #20

DATE: 05/11/98  
TIME: 13:49:07

**INPUT SET: S25700.raw**

**This Raw Listing contains the General Information Section and up to the first 5 pages.**

*raw*  
8/24/99

## **SEQUENCE LISTING**

ENTERED

3 (1) General Information:

4

5 (i) APPLICANT: Aldis Darzins  
6 Gregory T. Mrachko

7

8 (ii) TITLE OF INVENTION: A Sphingomonas Biodesulfurization  
9 Catalyst

10

11 (iii) NUMBER OF SEQUENCES: 13

12

13 (iv) CORRESPONDENCE ADDRESS:  
14 (A) ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.  
15 (B) STREET: Two Militia Drive  
16 (C) CITY: Lexington  
17 (D) STATE: Massachusetts  
18 (E) COUNTRY: USA  
19 (F) ZIP: 02173

20

21 (v) COMPUTER READABLE FORM:  
22 (A) MEDIUM TYPE: Floppy disk  
23 (B) COMPUTER: IBM PC compatible  
24 (C) OPERATING SYSTEM: PC-DOS/MS-DOS  
25 (D) SOFTWARE: PatentIn Release #1.0, Version #1.30

26

27 (vi) CURRENT APPLICATION DATA:  
28 (A) APPLICATION NUMBER: US 08/851,089  
29 (B) FILING DATE: 05-MAY-1997  
30 (C) CLASSIFICATION:

31

32 (vii) PRIOR APPLICATION DATA:  
33 (A) APPLICATION NUMBER: US 08/835,292  
34 (B) FILING DATE: 07-APR-1997

35

36 (viii) ATTORNEY/AGENT INFORMATION:  
37 (A) NAME: Elmore, Carolyn S.  
38 (B) REGISTRATION NUMBER: 37,567  
39 (C) REFERENCE/DOCKET NUMBER: EBC97-06A2

40

41 (ix) TELECOMMUNICATION INFORMATION:  
42 (A) TELEPHONE: (781) 861-6240  
43 (B) TELEFAX: (781) 861-9540

44

45

46 (2) INFORMATION FOR SEQ ID NO:1:

RAW SEQUENCE LISTING  
PATENT APPLICATION US/08/851,089DATE: 05/11/98  
TIME: 13:49:08

INPUT SET: S25700.raw

47  
48       (i) SEQUENCE CHARACTERISTICS:  
49           (A) LENGTH: 1362 base pairs  
50           (B) TYPE: nucleic acid  
51           (C) STRANDEDNESS: single  
52           (D) TOPOLOGY: linear

53  
54       (ii) MOLECULE TYPE: DNA (genomic)

55  
56  
57       (ix) FEATURE:  
58           (A) NAME/KEY: CDS  
59           (B) LOCATION: 1..1359

60  
61       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

62       ATG ACC GAT CCA CGT CAG CTG CAC CTG GCC GGA TTC TTC TGT GCC GGC      48  
63       Met Thr Asp Pro Arg Gln Leu His Leu Ala Gly Phe Phe Cys Ala Gly  
64       1               5               10               15  
65  
66       AAC GTC ACG CAC GCC CAC GGA GCG TGG CGC CAC GCC GAC GAC TCC AAC      96  
67       Asn Val Thr His Ala His Gly Ala Trp Arg His Ala Asp Asp Ser Asn  
68       20               25               30  
69  
70       GGC TTC CTC ACC AAG GAG TAC TAC CAG CAG ATT GCC CGC ACG CTC GAG      144  
71       Gly Phe Leu Thr Lys Glu Tyr Tyr Gln Gln Ile Ala Arg Thr Leu Glu  
72       35               40               45  
73  
74       CGC GGC AAG TTC GAC CTG CTG TTC CTT CCC GAC GCG CTC GCC GTG TGG      192  
75       Arg Gly Lys Phe Asp Leu Leu Phe Leu Pro Asp Ala Leu Ala Val Trp  
76       50               55               60  
77  
78       GAC AGC TAC GGC GAC AAT CTG GAG ACC GGT CTG CGG TAT GGC GGG CAA      240  
79       Asp Ser Tyr Gly Asp Asn Leu Glu Thr Gly Leu Arg Tyr Gly Gln  
80       65               70               75               80  
81  
82       GGC GCG GTG ATG CTG GAG CCC GGC GTA GTT ATC GCC GCG ATG GCC TCG      288  
83       Gly Ala Val Met Leu Glu Pro Gly Val Val Ile Ala Ala Met Ala Ser  
84       85               90               95  
85  
86       GTG ACC GAA CAT CTG GGG CTG GGC GCC ACC ATT TCC ACC ACC TAC TAC      336  
87       Val Thr Glu His Leu Gly Leu Gly Ala Thr Ile Ser Thr Thr Tyr Tyr  
88       100               105               110  
89  
90       CCG CCC TAC CAT GTA GCC CGG GTC GTC GCT TCG CTG GAC CAG CTG TCC      384  
91       Pro Pro Tyr His Val Ala Arg Val Val Ala Ser Leu Asp Gln Leu Ser  
92       115               120               125  
93  
94       TCC GGG CGA GTG TCG TGG AAC GTG GTC ACC TCG CTC AGC AAT GCA GAG      432  
95       Ser Gly Arg Val Ser Trp Asn Val Val Thr Ser Leu Ser Asn Ala Glu  
96       130               135               140  
97  
98  
99

**RAW SEQUENCE LISTING  
PATENT APPLICATION US/08/851,089**

DATE: 05/11/98  
TIME: 13:49:08

**INPUT SET: S25700.raw**

100	GCG CGC AAC TTC GGC TTC GAT GAA CAT CTC GAC CAC GAT GCC CGC TAC	480
101	Ala Arg Asn Phe Gly Phe Asp Glu His Leu Asp His Asp Ala Arg Tyr	
102	145 150	155 160
103		
104	GAT CGC GCC GAT GAA TTC CTC GAG GTC GTG CGC AAG CTC TGG AAC AGC	528
105	Asp Arg Ala Asp Glu Phe Leu Glu Val Val Arg Lys Leu Trp Asn Ser	
106	165	170 175
107		
108		
109		
110	TGG GAT CGC GAT GCG CTG ACA CTC GAC AAG GCA ACC GGC CAG TTC GCC	576
111	Trp Asp Arg Asp Ala Leu Thr Leu Asp Lys Ala Thr Gly Gln Phe Ala	
112	180	185 190
113		
114	GAT CCG GCT AAG GTG CGC TAC ATC GAC CAC CGC GGC GAA TGG CTC AAC	624
115	Asp Pro Ala Lys Val Arg Tyr Ile Asp His Arg Gly Glu Trp Leu Asn	
116	195	200 205
117		
118	GTA CGC GGG CCG CTT CAG GTG CCG CGC TCC CCC CAG GGC GAG CCT GTC	672
119	Val Arg Gly Pro Leu Gln Val Pro Arg Ser Pro Gln Gly Glu Pro Val	
120	210	215 220
121		
122	ATT CTG CAG GCC GGG CTT TCG GCG CGG GGC AAG CGC TTC GCC GGG CGC	720
123	Ile Leu Gln Ala Gly Leu Ser Ala Arg Gly Lys Arg Phe Ala Gly Arg	
124	225	230 235 240
125		
126	TGG GCG GAC GCG GTG TTC ACG ATT TCG CCC AAT CTG GAC ATC ATG CAG	768
127	Trp Ala Asp Ala Val Phe Thr Ile Ser Pro Asn Leu Asp Ile Met Gln	
128	245	250 255
129		
130	GCC ACG TAC CGC GAC ATA AAG GCG CAG GTC GAG GCC GCC GGA CGC GAT	816
131	Ala Thr Tyr Arg Asp Ile Lys Ala Gln Val Glu Ala Ala Gly Arg Asp	
132	260	265 270
133		
134	CCC GAG CAG GTC AAG GTG TTT GCC GCG GTG ATG CCG ATC CTC GGC GAG	864
135	Pro Glu Gln Val Lys Val Phe Ala Ala Val Met Pro Ile Leu Gly Glu	
136	275	280 285
137		
138	ACC GAG GCG ATC GCC AGG CAG CGT CTC GAA TAC ATA AAT TCG CTG GTG	912
139	Thr Glu Ala Ile Ala Arg Gln Arg Leu Glu Tyr Ile Asn Ser Leu Val	
140	290	295 300
141		
142	CAT CCC GAA GTC GGG CTT TCT ACG TTG TCC AGC CAT GTC GGG GTC AAC	960
143	His Pro Glu Val Gly Leu Ser Thr Leu Ser Ser His Val Gly Val Asn	
144	305	310 315 320
145		
146	CTT GCC GAC TAT TCG CTC GAT ACC CCG CTG ACC GAG GTC CTG GGC GAT	1008
147	Leu Ala Asp Tyr Ser Leu Asp Thr Pro Leu Thr Glu Val Leu Gly Asp	
148	325	330 335
149		
150	CTC GCC CAG CGC AAC GTG CCC ACC CAA CTG GGC ATG TTC GCC AGG ATG	1056
151	Leu Ala Gln Arg Asn Val Pro Thr Gln Leu Gly Met Phe Ala Arg Met	
152	340	345 350

RAW SEQUENCE LISTING  
PATENT APPLICATION US/08/851,089DATE: 05/11/98  
TIME: 13:49:09

INPUT SET: S25700.raw

153  
154 TTG CAG GCC GAG ACG CTG ACC GTG GGA GAA ATG GGC CGG CGT TAT GGC 1104  
155 Leu Gln Ala Glu Thr Leu Thr Val Gly Glu Met Gly Arg Arg Tyr Gly  
156 355 360 365  
157  
158 GCC AAC GTG GGC TTC GTC CCG CAG TGG GCG GGA ACC CGC GAG CAG ATC 1152  
159 Ala Asn Val Gly Phe Val Pro Gln Trp Ala Gly Thr Arg Glu Gln Ile  
160 370 375 380  
161  
162  
163  
164 GCG GAC CTG ATC GAG ATC CAT TTC AAG GCC GGC GCC GAT GGC TTC 1200  
165 Ala Asp Leu Ile Glu Ile His Phe Lys Ala Gly Gly Ala Asp Gly Phe  
166 385 390 395 400  
167  
168 ATC ATC TCG CCG GCG TTC CTG CCC GGA TCT TAC GAG GAA TTC GTC GAT 1248  
169 Ile Ile Ser Pro Ala Phe Leu Pro Gly Ser Tyr Glu Glu Phe Val Asp  
170 405 410 415  
171  
172 CAG GTG GTG CCC ATC CTG CAG CAC CGC GGA CTG TTC CGC ACT GAT TAC 1296  
173 Gln Val Val Pro Ile Leu Gln His Arg Gly Leu Phe Arg Thr Asp Tyr  
174 420 425 430  
175  
176 GAA GGC CGC ACC CTG CGC AGC CAT CTG GGA CTG CGT GAA CCC GCA TAC 1344  
177 Glu Gly Arg Thr Leu Arg Ser His Leu Gly Leu Arg Glu Pro Ala Tyr  
178 435 440 445  
179  
180 CTG GGA GAG TAC GCA TGA 1362  
181 Leu Gly Glu Tyr Ala  
182 450  
183  
184  
185 (2) INFORMATION FOR SEQ ID NO:2:  
186  
187 (i) SEQUENCE CHARACTERISTICS:  
188 (A) LENGTH: 453 amino acids  
189 (B) TYPE: amino acid  
190 (D) TOPOLOGY: linear  
191  
192 (ii) MOLECULE TYPE: protein  
193  
194 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:  
195  
196 Met Thr Asp Pro Arg Gln Leu His Leu Ala Gly Phe Phe Cys Ala Gly  
197 1 5 10 15  
198  
199 Asn Val Thr His Ala His Gly Ala Trp Arg His Ala Asp Asp Ser Asn  
200 20 25 30  
201  
202 Gly Phe Leu Thr Lys Glu Tyr Tyr Gln Gln Ile Ala Arg Thr Leu Glu  
203 35 40 45  
204  
205 Arg Gly Lys Phe Asp Leu Leu Phe Leu Pro Asp Ala Leu Ala Val Trp

**RAW SEQUENCE LISTING  
PATENT APPLICATION US/08/851,089**

DATE: 05/11/98  
TIME: 13:49:10

### **INPUT SET: S25700.raw**

206	50	55	60
207			
208	Asp Ser Tyr Gly Asp Asn Leu Glu Thr Gly Leu Arg Tyr Gly Gly Gln		
209	65	70	75
210			80
211	Gly Ala Val Met Leu Glu Pro Gly Val Val Ile Ala Ala Met Ala Ser		
212	85	90	95
213			
214	Val Thr Glu His Leu Gly Leu Gly Ala Thr Ile Ser Thr Thr Tyr Tyr		
215	100	105	110
216			
217			
218	Pro Pro Tyr His Val Ala Arg Val Val Ala Ser Leu Asp Gln Leu Ser		
219	115	120	125
220			
221	Ser Gly Arg Val Ser Trp Asn Val Val Thr Ser Leu Ser Asn Ala Glu		
222	130	135	140
223			
224	Ala Arg Asn Phe Gly Phe Asp Glu His Leu Asp His Asp Ala Arg Tyr		
225	145	150	155
226			160
227	Asp Arg Ala Asp Glu Phe Leu Glu Val Val Arg Lys Leu Trp Asn Ser		
228	165	170	175
229			
230	Trp Asp Arg Asp Ala Leu Thr Leu Asp Lys Ala Thr Gly Gln Phe Ala		
231	180	185	190
232			
233	Asp Pro Ala Lys Val Arg Tyr Ile Asp His Arg Gly Glu Trp Leu Asn		
234	195	200	205
235			
236	Val Arg Gly Pro Leu Gln Val Pro Arg Ser Pro Gln Gly Glu Pro Val		
237	210	215	220
238			
239	Ile Leu Gln Ala Gly Leu Ser Ala Arg Gly Lys Arg Phe Ala Gly Arg		
240	225	230	235
241			240
242	Trp Ala Asp Ala Val Phe Thr Ile Ser Pro Asn Leu Asp Ile Met Gln		
243	245	250	255
244			
245	Ala Thr Tyr Arg Asp Ile Lys Ala Gln Val Glu Ala Ala Gly Arg Asp		
246	260	265	270
247			
248	Pro Glu Gln Val Lys Val Phe Ala Ala Val Met Pro Ile Leu Gly Glu		
249	275	280	285
250			
251	Thr Glu Ala Ile Ala Arg Gln Arg Leu Glu Tyr Ile Asn Ser Leu Val		
252	290	295	300
253			
254	His Pro Glu Val Gly Leu Ser Thr Leu Ser Ser His Val Gly Val Asn		
255	305	310	315
256			320
257	Leu Ala Asp Tyr Ser Leu Asp Thr Pro Leu Thr Glu Val Leu Gly Asp		
258	325	330	335

PAGE: 1

**SEQUENCE VERIFICATION REPORT**  
PATENT APPLICATION **US/08/851,089**

DATE: 05/11/98  
TIME: 13:49:11

***INPUT SET: S25700.raw***

Line

Error

Original Text